Multi-Node RSS-based Localization with the Aid of Compressed Sensing: An ℓ_1-localization Approach

Ehsan Zandi, and Rudolf Mathar
Institute for Theoretical Information Technology, RWTH Aachen University, Germany
Email: {zandi, mathar}@ti.rwth-aachen.de

***Abstract*—In this work try we try to estimate the positions of multiple co-channel wireless nodes along with the unknown transmit power of them. The propagation channel is assumed to be log-normal shadowing model. We propose an unbiased estimator. The underlying complicated optimization problem has a combinatorial nature that selects the best grid points as the location of the targets. We then convert the combinatorial problem to a convex form by means of ℓ_1-minimization, or precisely a technique which is inspired by the theory of *compressed sensing* (CS). The performance of the estimator is justified to be good using simulations.**

***Index Terms*—multi-source localization, mixed-integer programming, k-NN, internet of things**

I. Introduction

It is envisaged that the majority of applications in the context of the internet of things and 5G mobile networks depend on the location awareness to deliver better services. Therefore, the old topic of localization is not yet obsolete. In the literature, a variety of techniques have been exploited to solve such a problem. We here stick to a *received signal strength* (RSS)-based technique due to its simplicity and lower cost compared to the *time difference of arrival* (TDoA) or *angle of arrival* (AoA), [1]. Despite its vulnerability against uncertainties of path-loss model, RSS localization is beneficial whenever the precision can be somewhat compromised for price. The RSS-based localization for a single target with unknown transmit power is studied in many publications, such as [2], where the by dividing the RSS of two different receivers the transmit power cancels out. The remaining of the problem is a standard multilateration problem. This technique is known as differential or ratio of RSS which is not applicable in our case since we assume that there are more than one transmitter on the same channel which causes co-channel interference problem.

A. Sparsity Aware Localization

Employing CS in the context of localization, especially in fingerprinting, is a famous technique. This is motivated by the spatial sparsity of target(s) transmitting in the area. Usually, a grid is defined on the area of interest, and then the RSS of the receivers, i.e., *Access Point*s (APs) are measured by moving the transmitter from one *grid point* (GP) to another in order to build a radio map. Later, based on the RSS observations, the GPs near which the targets reside will be identified by finding a vector whose ℓ_0-norm is equal to the number of targets. That means a vector whose all entries are 0, except for those entries which point out to the existence of a target. Since ℓ_0-norm is combinatorial and non-convex, the vector can be reconstructed using CS, i.e., ℓ_1-minimization, with a high probability if some certain conditions are met. One of these conditions is sparsity which is satisfied, since out of many GPs only few may host a target.

Among the works that do multi-target CS-fingerprinting, we can mention [3], [4]. The difference between their work and ours is that they assume for each target there is a distinct observation (RSS) at *fusion center* (FC). Indeed, in their scenario, the targets are not co-channel. This can be seen as several single target problems.

On the contrary, [5] applies CS for the co-channel multi-target fingerprinting scenario, where not only the RSS at the receivers but also the cross-correlation between signals of the target signals are exploited for improving the quality of the localization. This, nevertheless, burdens the communication between the APs and FC leading to higher power consumption as well as implementation costs, cf. [6], [7]. Therefore, it is not a reasonable strategy in the case of a sensor network, where the battery consumption of *sensor node*s (SNs) is a limiting factor. Furthermore, [8] and [9] tackle the problem of co-channel multi-target localization. While the former does so by applying CS to a TDoA system, the latter formulates an AoA-based ℓ_0-norm selection problem to decide which peak in the auto-correlation function of the receiver belongs to which transmitter. Even though authors do not use the term *compressed sensing*, yet their solution is based on ℓ_1-minimization.

B. Our Contribution

In this work, we assume a log-normal shadowing path-loss model, where multiple co-channel transmitters cause interference one another. This makes the multilateration technique, as in the single target scenario, impossible. To the best of our knowledge, there is no work with similar assumptions, except for our previous works [6] and [7], where ℓ_0-minimization techniques are deployed to solve the intractable underlying mathematical problem. In this work, we transform the problem to a sparse formulation and solve an ℓ_1-minimization. Note our solution is inspired by the theory of CS in the sense that we relax ℓ_0-norm to ℓ_1-norm. But our formulation does not have the exact form of the famous techniques of CS. Therefore, we simply call the method of this paper ℓ_1-localization. We start with the assumption that targets are at grid points and deal with the off grid targets, subsequently. An interesting application for our scenario is finding the position of the illegitimate secondary user(s) with unknown transmit power. Note such interfering users decrease the throughput of the primary user or even cause link failure due to strong interference.

The organization of this paper is as follows: the system model is described in Sec. II and statistical properties of the RSS at receivers in Sec. III. Then, Sec. IV explains the idea of ℓ_1-localization and also presents the proposed algorithm. The performance of the presented solutions is justified by means of computer simulations in Sec. V.

Notations: All mathematical notations, symbols and variables of this paper are summarized in Tab. I.

II. System Model

The system of consideration consists of $N \in \mathbb{N}$ active targets with unknown position and $K \in \mathbb{N}$ passive SNs with known positions. Each target $n \in \mathbb{I}_N$ transmits a signal with the unknown power p_n. We know that that transmit power of each target is bounded as follows

$$\underline{P} \leq p_n \leq \bar{P}\,,\ \forall n \in \mathbb{I}_N\,, \tag{1}$$

where $\underline{P}$, $\bar{P} \in \mathbb{R}_+$ are, respectively, the lowest and highest possible values for the transmit power of an active target. The propagation channel is based on the log-normal shadowing attenuation model presented in [10]. In a multi-source scenario, the RSS r_k at sensor k is the sum of different terms corresponding to the received power of each target signal [11], [12]:

$$r_k = \sum_{n\in\mathbb{I}_N} c_0\, p_n\, d_{kn}^{-\alpha}\, 10^{\frac{\zeta_{kn}}{10}}\,, \tag{2}$$

where d_{kn} is the distance between sensor k and n^{th} target, α is the path-loss exponent and $\zeta_{kn} \sim \mathcal{N}(0, \sigma_{kn}^2)$ is a zero-mean Gaussian random variable with the power of σ_{kn}^2. It models the log-normal shadowing between each pair of sensor and target

TABLE I: Summary of general mathematical notations

Notation	Description
$\mathbb{N}$	set of all integer positive and non-zero numbers
$\mathbb{R}$	set of all real numbers
$\mathbb{R}_+$	set of all non-negative real numbers
$\mathbb{R}^{m\times n}$	set of all real matrices of size of $m \times n$
$\mathbb{R}^m$	set of all real vectors of size of $m \times 1$
$\mathbb{I}_l$	the index-set $\mathbb{I}_l := \{1, \cdots, l\}$ for $l \in \mathbb{N}$
x	scalar x
$\mathbf{x}$	column vector $\mathbf{x}$ with entries x_i
$\mathbf{x}'$	transpose of vector $\mathbf{x}$
$\mathbf{X}$	matrix $\mathbf{X}$ with entries x_{ij} or $[\mathbf{X}]_{ij}$
$[\mathbf{x}]_i$ or x_i	entry i of vector $\mathbf{x}$
$[\mathbf{X}]_{ij}$ or x_{ij}	entry i, j of matrix $\mathbf{X}$
$\mathbf{1}$	all-ones vector of proper size
$\Vert\mathbf{x}\Vert_0$	ℓ_0-norm, i.e., the number of non-zero entries of $\mathbf{x}$
$(\cdot)^\star$	optimal solution of an optimization problem

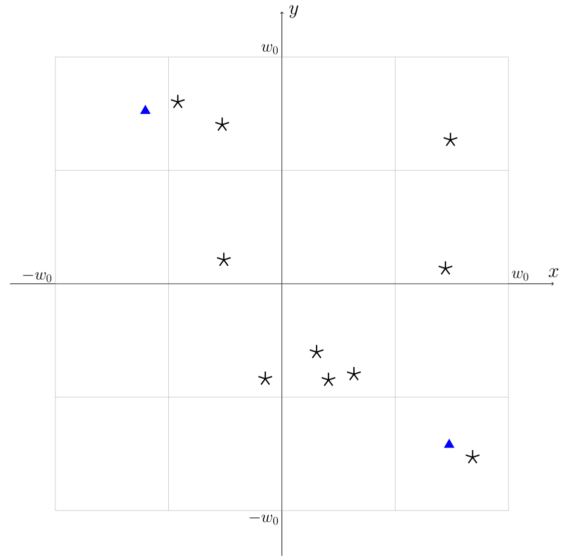

Fig. 1: A wireless sensor network consisting of $K = 10$ passive sensors ($\star$) and $N = 2$ targets (▲) . The grid granularity is $G = 5$, which means the area of interest is divided to $(G-1)^2 = 16$ smaller squares. This leads to $G^2 = 25$ GPs, around which we look for the targets. The width of each grid square is $\Delta_g = \frac{w}{2}$.

nodes and is assumed to be *identically and independently distributed* (iid). The coefficient c_0 is given by [10],

$$c_0 := \frac{G_t G_r \lambda^2}{(4\pi)^2}\,, \tag{3}$$

where G_t and G_r are the gains of transmit and receive antennae, respectively. The wavelength is denoted by λ. We assume that c_0 is known and without loss of generality and for the sake of simplicity, $c_0 = 1$.

Here, we have neglected the thermal additive noise due to the fact that shadowing has a much stronger effect on RSS compared to the thermal noise [13], [14], [15]. The main reason for such an assumption is that the receivers have considerably higher detection threshold than sensitivity. Furthermore, the RSS measurements are usually performed after correctly decoding of the information data out of the received data packets [13]. Besides, the effect of additive noise can be somewhat compensated using methods of blind estimation of noise power, e.g., [16], [17].

The area of observation is assumed to be a square in the range of $[-w, w]$, $w \in \mathbb{R}_+$ in both x- and y- axes, in the Cartesian coordinate system. The targets and sensors are randomly distributed within the area. The ordered pair $(\check{x}_k, \check{y}_k)$ stands for the coordinate of k^{th} sensor node, while target n is located at the unknown position (x_n, y_n). Assuming that FC acquires the values of RSS r_k of the k^{th} sensor error-freely upon successful communication from SN, it has to solve the following system of nonlinear and non-convex equation to find the position (x_n, y_n) of each target:

$$r_k = \sum_{n \in \mathbb{I}_N} \frac{p_n \, 10^{\frac{\zeta_{kn}}{10}}}{\left(\sqrt{(x_n - \check{x}_k)^2 + (y_n - \check{y}_k)^2}\right)^{\alpha}} \, . \quad (4)$$

Unfortunately, solving such a system of equations is not easy. Thus, we resort to solve it by discretizing the area into a grid of granularity of $G \in \mathbb{N}$ which means G^2 GPs, in total. Let $\mathcal{G}_G^w(x, y)$ define the grid set centered at the point (x, y) of width $2w \in \mathbb{R}_+$ and the granularity G. Indeed, it is the set of the equidistant GPs defined by

$$\mathcal{G}_G^w(x, y) := \left\{(x - w + (i-1)\Delta_g, y - w + (j-1)\Delta_g) \,|\, i, j \in \mathbb{I}_G\right\}, \quad (5)$$

where $\Delta_g = \frac{2w}{G-1}$ is the width of one grid square. Then, our defined grid consists of the GPs $(\tilde{x}_m, \tilde{y}_m) \in \mathcal{G}_G^w(0, 0)$, $m \in \mathbb{I}_{G^2}$. Fig. 1 depicts an example grid given by $\mathcal{G}_5^w(0, 0)$.

III. Sum of Log Normal Random Variables

The sum of *log-normal* ($\mathcal{LN}$) random variables has an unknown *probability distribution function* (pdf) [18], even for the sum of two random variables. The underlying reason is that the moment generating function of the $\mathcal{LN}$ distribution is not defined, [19]. In general, $\mathcal{LN}$ distribution has bad behavior and cannot be described by its moments. This holds, consequently, for *sum log-normal* ($\mathcal{SLN}$) distribution. On the other hand, all the moments of the $\mathcal{LN}$ distribution exist in closed-form. In the literature, there are several works trying to approximate $\mathcal{SLN}$, cf. [20] and references therein, based on either $\mathcal{LN}$ approximation, e.g., [18], or numerical methods. Since our goal is to come up with a good location estimator, we are interested in a closed-form approximation which does not need prior knowledge of the *cumulative distribution function* (cdf) of the RSS at each sensor. Therefore, we choose the Fenton-Wilkinson method in [18], which approximates $\mathcal{SLN}$ by $\mathcal{LN}$ and matching the first and second moments. Though, this is not very accurate in low values of the random variable, but is suitable for our purpose due to its simplicity and due to the fact the all the moments of $\mathcal{SLN}$ distribution exist in closed-form. We will see in what follows, that even this simple approximation leads to a very complicated optimization problem for finding the position of the targets.

Let R_{kn} be the random variable from which the values of r_{kn} are drawn. Note that r_{kn} stands for the RSS at sensor k due to the transmit signal from n^{th} target, i.e, $r_{kn} = p_n \, d_{kn}^{-\alpha} \, 10^{\frac{\zeta_{kn}}{10}}$, then the l^{th} moment of R_{kn} is given by

$$\mathcal{E}\left(R_{kn}^l\right) = \left(p_n \, d_{kn}^{-\alpha}\right)^l \beta_{kn}^{l^2}, \quad (6)$$

where $\beta_{kn} = e^{\frac{(\ln 10)^2 \sigma_{kn}^2}{200}}$. Obviously, the mean and variance of R_{kn} read

$$\mathcal{E}(R_{kn}) = p_n \, d_{kn}^{-\alpha} \beta_{kn} \, , \quad (7)$$

$$\mathrm{Var}(R_{kn}) = (p_n \, d_{kn}^{-\alpha})^2 (\beta_{kn}^2 - 1) \beta_{kn}^2 \, . \quad (8)$$

Let us assume that all σ_{kn} are equal, then define $\beta = \beta_{kn}$ for all pair of sensors and targets. A scenario that such an assumption does not hold is multi-floor indoor environments [11]. We further assume that the value of β is known. Since all the random variables R_{kn} are pairwise independent the mean M_k and variance V_k of the random variable $R_k := \sum_{n \in \mathbb{I}_N} R_{kn}$ is given by

$$M_k = \beta g_k \, , \; g_k := \sum_{n \in \mathbb{I}_N} p_n \, d_{kn}^{-\alpha} \, , \quad (9a)$$

$$V_k = (\beta^2 - 1)\beta^2 h_k \, , \; h_k := \sum_{n \in \mathbb{I}_N} (p_n \, d_{kn}^{-\alpha})^2 \, . \quad (9b)$$

Now, similar to [18] we approximate the random variable R_k by an $\mathcal{LN}$:

$$R_k \approx e^{\mu_k + \sigma_k X} \, , \quad (10)$$

where X is a zero-mean normal random variable with variance one. We find the values of μ_k and σ_k such that the mean and the variance of R_k equate with the ones of the random variable $e^{\mu_k + \sigma_k X}$:

$$M_k = e^{\mu_k + \frac{\sigma_k^2}{2}} \, , \quad (11)$$

$$V_k = e^{2\mu_k + 2\sigma_k^2} \, , \quad (12)$$

which results in

$$\mu_k = 2\ln(M_k) - \frac{1}{2}\ln(M_k^2 + V_k) \, , \quad (13a)$$

$$\sigma_k^2 = \ln(M_k^2 + V_k) - 2\ln M_k \, . \quad (13b)$$

Finally, we state that the $\ln r_k$ is approximately a normal random variable:

$$\ln r_k \approx \mu_k + \sigma_k X \, . \quad (14)$$

Deriving classical estimators, e.g., *maximum likelihood* (ML) or *minimum variance unbiased* (MVU) is not possible, since the exact pdf of the RSS is unknown. We are rather interested in an estimator which is only unbiased since minimizing the *mean square error* (MSE) w.r.t the given approximation, i.e., minimizing σ_k^2, is also mathematically intractable.

IV. ℓ_1-Localization

Inspired by the idea of employing CS in the context of fingerprinting localization, we try to take advantage of ℓ_1-minimization in our problem. Since we exploit the explicit path-loss model we can arbitrarily increase the granularity of the (virtual) radio map.

A. On-grid Target

Let us momentarily assume that the targets are located exactly at GPs. We later deal with the case that such an assumption is invalid. Let $\tilde{d}_{km}$ denote the distance between sensor $k \in \mathbb{I}_k$ and the m^{th} GP

$$\tilde{d}_{km} = \sqrt{(\tilde{x}_m - \check{x}_k)^2 + (\tilde{y}_m - \check{y}_k)^2} \, , \; m \in \mathbb{I}_M \, . \quad (15)$$

which results in RSS $\tilde{r}_k$ at the SN k:

$$\tilde{r}_k = \sum_{m \in \mathbb{I}_M} \tilde{p}_m \tilde{d}_{km}^{-\alpha} 10^{\frac{\zeta_{kn}}{10}} \, . \quad (16)$$

where the variable $\tilde{p}_m$ is zero when the GP is not occupied by any target. It is, otherwise, equal to the transmit power of the corresponding target, i.e.,

$$\tilde{p}_m = \begin{cases} 0\,, & \nexists\, n \in \mathbb{I}_N \mid (x_n, y_n) = (\tilde{x}_m, \tilde{y}_m) \\ p_n\,, & \exists\, n \in \mathbb{I}_N \mid (x_n, y_n) = (\tilde{x}_m, \tilde{y}_m) \end{cases} . \tag{17}$$

Therefore, we can write the RSS readings of all sensors as $\tilde{\mathbf{r}} \in \mathbb{R}_+^K$ by

$$\tilde{\mathbf{r}} = \tilde{\mathbf{\Phi}}\tilde{\mathbf{p}}\,, \tag{18}$$

where $[\tilde{\mathbf{\Phi}}]_{km} := \tilde{d}_{km}^{-\alpha}$ and the k^{th} entry of $\tilde{\mathbf{r}}$ is $\tilde{r}_k$. The m^{th} entry of $\tilde{\mathbf{p}} \in \mathbb{R}^M$, i.e., $\tilde{p}_m \in \{0, [\underline{P}, \bar{P}]\}$ is a selection variable to choose (or to not choose) the m^{th} GP. Since exactly N GPs out of M must be selected, the vector $\tilde{\mathbf{p}}$ is N-sparse.

To find the position and the transmit power of the targets we minimize $\|\mathbf{r} - \tilde{\mathbf{r}}\|_2$, where the entries of the vector $\mathbf{r} \in \mathbb{R}_+^K$ are the RSS-readings at SNs. Hence, it is desirable to solve the optimization problem

$$\min_{\tilde{\mathbf{p}}} \;\; \|\mathbf{r} - \tilde{\mathbf{\Phi}}\tilde{\mathbf{p}}\|_2 \tag{19a}$$

$$\text{s.t.} \;\; \|\tilde{\mathbf{p}}\|_0 = N \tag{19b}$$

$$\mathbf{1}\underline{P} \leq \tilde{\mathbf{p}} \leq \mathbf{1}\bar{P}\,. \tag{19c}$$

Unfortunately, due to the constraint (19c) the problem does not have exactly the form of either *Basis Pursuit* (BP), *Basis Pursuit Denoising Sensing* (BPDN) or *Dantzig Selector* (DS). Therefore, we cannot directly apply the CS theory to our problem. Consequently, we propose a different solution which is simply based on the idea of relaxing the ℓ_0-norm to ℓ_1-norm. Therefore, we call what comes in the following ℓ_1-localization. But before doing so we need to do a minor modification to the equation (18):

$$\tilde{\mathbf{r}} = \mathbf{\Phi}\mathbf{s}\,, \tag{20}$$

where $[\mathbf{\Phi}]_{km} := \tilde{p}_m \tilde{d}_{km}^{-\alpha}$ and $\mathbf{s} \in \mathbb{R}^M$. Indeed, s_m is a selection variable to choose (or to not choose) the m^{th} GP, which means $s_m \in \{0, 1\}$. Thus, $\mathbf{s}$ must be a N-sparse vector. Since the transmit power $\tilde{p}_m$ is unknown, we assume

$$\phi_{km} = [\mathbf{\Phi}]_{km} = \frac{1}{2}(\underline{P} + \bar{P})\tilde{d}_{km}^{-\alpha}\,. \tag{21}$$

In case the transmit power is known, i.e., $\underline{P} = \bar{P}$, the aforementioned assumption is correct because $\tilde{p}_m = \frac{1}{2}(\underline{P} + \bar{P}) = \underline{P} = \bar{P}$. The idea of ℓ_1-localization for the on-grid targets is based on relaxing the following combinatorial problem

$$\min_{\mathbf{s}} \;\; \|\mathbf{r} - \mathbf{\Phi}\mathbf{s}\|_2 \tag{22a}$$

$$\text{s.t.} \;\; \mathbf{s} \in \{0, 1\}^M \tag{22b}$$

$$\|\mathbf{s}\|_0 = N\,, \tag{22c}$$

to the convex *quadratic program* (QP), below

$$\min_{\mathbf{s}} \;\; \|\mathbf{\Psi}\mathbf{r} - \mathbf{Q}\mathbf{s}\|_2 \tag{23a}$$

$$\text{s.t.} \;\; \mathbf{0} \leq \mathbf{s} \leq \mathbf{1} \tag{23b}$$

$$\mathbf{1}'\mathbf{s} = N\,, \tag{23c}$$

where $\mathbf{Q} := \mathbf{\Psi}\mathbf{\Phi}$ and $\mathbf{\Psi} = orth(\mathbf{\Phi}')'\mathbf{\Phi}^\dagger$ is the pre-processing matrix of size $K \times K$. The symbol † shows the Moore–Penrose inverse and $orth(\mathbf{X})$ is an orthogonal basis for the range of matrix $\mathbf{X}$. The authors in [3] apply such a pre-processing by multiplying both sides of the equation $\mathbf{r} = \mathbf{\Phi}\mathbf{s}$ with $\mathbf{\Psi}$, since the sensing matrix $\mathbf{\Phi}$ does not possess the incoherence property.

The problem (23) can be solved by any of the methods *simplex algorithm*, *interior point*, *augmented Lagrangian* and *gradient descent*. But as our focus in this work is not numerical aspects of optimization, we rather solve the problem by existing solvers such as [21] or [22].

1) Averaging: Since the targets can be anywhere within the area of observation, the optimal solution of the problem (23) does not provide a good solution. The reason is simply that the optimal point $\mathbf{s}^\star$ is not necessarily N-sparse. Therefore, we apply the averaging rule for a better position estimation

$$\hat{x}_n = \frac{\sum\limits_{m \in \Pi_n} s_m^\star x_m}{\sum\limits_{m \in \Pi_n} s_m^\star}\,, \tag{24a}$$

$$\hat{y}_n = \frac{\sum\limits_{m \in \Pi_n} s_m^\star y_m}{\sum\limits_{m \in \Pi_n} s_m^\star}\,, \tag{24b}$$

where the sets $\Pi_n \subset \bar{\mathbb{I}}_{M_0} \subset \mathbb{I}_M$, $\forall n \in \mathbb{I}_N$ are the partitions of the set $\bar{\mathbb{I}}_{M_0}$, i.e.,

$$\Pi_n \cap \Pi_{n'} = \emptyset\,,\; \forall n, n' \in \bar{\mathbb{I}}_{M_0}\; n \neq n'\,, \tag{25a}$$

$$\Pi_1 \cup \cdots \cup \Pi_N = \bar{\mathbb{I}}_{M_0}\,. \tag{25b}$$

The set $\bar{\mathbb{I}}_{M_0}$ is the index set of M_0 largest entries of $\mathbf{s}^\star$, for a given $M_0 < M$.

2) Clustering the GPs: By solving the problem (23), the set $\bar{\mathbb{I}}_{M_0}$ is easily identified for a given M_0. Now the important question arises how to decide on the optimal partitioning. We exploit the technique of k-means clustering where M_0 observations will be clustered into N partitions. $\Pi_1, \cdots, \Pi_N$.

Fig. 2 shows how the idea averaging using (24) improves the positioning performance. Note, if we choose $M_0 = N$ it means we take the N largest entries of $\mathbf{s}^\star$ and select the corresponding GPs as the position estimation.

3) Iterative grid refinement: Assume starting from an initial set $\mathcal{G}_0$ of GPs and applying the aforementioned technique for localization, then we can define new grid around the points $(\hat{x}_n, \hat{y}_n)$ and repeat the whole process. Given the set of points $(\hat{x}_n, \hat{y}_n)$, the new grid will be formed by

$$\mathcal{G}(N) = \bigcup_{n \in \mathbb{I}_N} \mathcal{G}_G^{\frac{w}{2}}(\hat{x}_n, \hat{y}_n)\,, \tag{26}$$

where $2w$ is the width of the monitoring area. Having updated the grid $\mathcal{G}(N)$, at iteration i the new position estimate $(\hat{x}_n, \hat{y}_n)$ is then achieved using (23) and (24). The simulation reveals that this idea provides a high convergence probability. Though, we observe ripples in the position estimations in the vicinity of the true position of targets. Therefore, after $I_1 \in \mathbb{N}$ iterations,

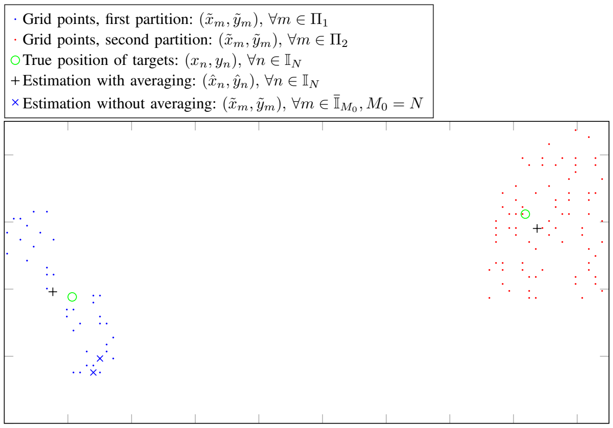

Fig. 2: The result of ℓ_1-localization via the problem (23) and then applying the averaging rule (24) for $N = 2$ targets.

we take the position estimation $(\hat{x}_n, \hat{y}_n)$ values as the initial points for another algorithm which we call *fine-tuning*.

4) Fine-tuning the grid: Similar to works [7] we are interested in solving the system of equations $\ln r_k - \mu_k = 0$, see Eq. (14). Since it is mathematically intractable to solve this system of equations, we minimize the function

$$\sum_{k \in \mathbb{I}_K} \tilde{f}_k^2 \tag{27}$$

where $\tilde{f}_k$ is the first order Taylor approximation of the function $\ln r_k - \mu_k$. Let us start doing this by defining the function $f_k(\tilde{p}_1, \cdots, \tilde{p}_M, \tilde{x}_1, \cdots, \tilde{x}_M, \tilde{y}_1, \cdots, \tilde{y}_M)$ or in short form f_k,

$$\begin{aligned} f_k := \ln r_k - \tilde{\mu}_k = \ln r_k - \ln \beta - 2 \ln \tilde{g}_k \\ + \frac{1}{2} \ln \left(\tilde{g}_k^2 + (\beta^2 - 1) \tilde{h}_k \right) , \end{aligned} \tag{28}$$

where

$$\tilde{g}_k = \sum_{m \in \mathbb{I}_M} \tilde{p}_m \, \tilde{d}_{km}^{-\alpha} , \tag{29}$$

$$\tilde{h}_k = \sum_{m \in \mathbb{I}_M} \left(\tilde{p}_m \, \tilde{d}_{km}^{-\alpha} \right)^2 . \tag{30}$$

Then, the derivative of f_k w.r.t θ that stands for any of the variables $\tilde{x}_m$, $\tilde{y}_m$, or $\tilde{p}_m$ reads

$$\frac{\partial f_k}{\partial \theta} = - \frac{2}{g_k} \frac{\partial g_k}{\partial \theta} + \frac{\frac{\partial g_k}{\partial \theta} g_k + \frac{1}{2}(\beta^2 - 1) \frac{\partial h_k}{\partial \theta}}{g_k^2 + (\beta^2 - 1) h_k} ,$$

where

$$\frac{\partial g_k}{\partial \tilde{x}_m} = \frac{\alpha \, \tilde{p}_m \, (\check{x}_k - \tilde{x}_m)}{\tilde{d}_{km}^{\alpha+2}}, \frac{\partial h_k}{\partial \tilde{x}_m} = \frac{2\alpha \, \tilde{p}_m^2 \, (\check{x}_k - \tilde{x}_m)}{\tilde{d}_{km}^{2\alpha+2}},$$

$$\frac{\partial g_k}{\partial \tilde{y}_m} = \frac{\alpha \, \tilde{p}_m \, (\check{y}_k - \tilde{y}_m)}{\tilde{d}_{km}^{\alpha+2}}, \frac{\partial h_k}{\partial \tilde{y}_m} = \frac{2\alpha \, \tilde{p}_m^2 \, (\check{y}_k - \tilde{y}_m)}{\tilde{d}_{km}^{2\alpha+2}},$$

$$\frac{\partial g_k}{\partial \tilde{p}_m} = \tilde{d}_{km}^{-\alpha} , \frac{\partial h_k}{\partial \tilde{p}_m} = 2 \, \tilde{p}_m \, \tilde{d}_{km}^{-2\alpha} .$$

Assume at the iteration i the vector $\boldsymbol{\theta}$ is given by $\boldsymbol{\theta} := (\tilde{p}_1^{i-1}, \cdots, \tilde{p}_M^{i-1}, \tilde{x}_1^{i-1}, \cdots, \tilde{x}_M^{i-1}, \tilde{y}_1^{i-1}, \cdots, \tilde{y}_M^{i-1})$. Furthermore, let us define a_{km}^{i-1}, b_{km}^{i-1}, and c_{km}^{i-1} as, respectively, $\frac{\partial f_k}{\partial \tilde{x}_m}(\boldsymbol{\theta})$, $\frac{\partial f_k}{\partial \tilde{y}_m}(\boldsymbol{\theta})$, $\frac{\partial f_k}{\partial \tilde{p}_m}(\boldsymbol{\theta})$, Thus, f_k at the i^{th} iteration can be approximated by its first order Taylor series expansion:

$$f_k \approx f_k^{i-1} + \sum_{m \in \mathbb{I}_M} a_{km}^{i-1} \mathrm{d}\tilde{x}_m + b_{km}^{i-1} \mathrm{d}\tilde{y}_m + c_{km}^{i-1} \mathrm{d}\tilde{p}_m ,$$

where $f_k^{i-1} := f(\boldsymbol{\theta})$, and $d\tilde{x}_m$, $d\tilde{y}_m$ and $d\tilde{p}_m$ are our optimization variables. In fact, from iteration $I_1 + 1$ onward we solve (31)

$$\min_{\substack{\mathrm{d}\tilde{x}_m, \mathrm{d}\tilde{y}_m, \\ \mathrm{d}\tilde{p}_m, m \in \mathbb{I}_M}} \quad \sum_{k \in \mathbb{I}_K} (f_k^{i-1} + \sum_{m \in \mathbb{I}_M} a_{km}^{i-1} \mathrm{d}\tilde{x}_m + b_{km}^{i-1} \mathrm{d}\tilde{y}_m + c_{km}^{i-1} \mathrm{d}\tilde{p}_m)^2 \tag{31a}$$

$$\text{s.t.} \quad \mathrm{d}\tilde{x}_m, \mathrm{d}\tilde{y}_m, \mathrm{d}\tilde{p}_m \in \mathbb{R} , \tag{31b}$$

$$\underline{P} - \tilde{p}_m^{i-1} \leq \mathrm{d}\tilde{p}_m \leq \bar{P} - \tilde{p}_m^{i-1} , \tag{31c}$$

$$- \delta \leq \mathrm{d}\tilde{x}_m \leq \delta , \tag{31d}$$

$$- \delta \leq \mathrm{d}\tilde{y}_m \leq \delta , \tag{31e}$$

for a grid of $M = N$ points given by

$$\mathcal{G}(N) = \bigcup_{n \in \mathbb{I}_N} \{(\hat{x}_n, \hat{y}_n)\} . \tag{32}$$

We solve this problem from iteration $i = I_1 + 1$ to $I_1 + I_2$ and update the estimate of position and transmit power by the following rule

$$\tilde{x}_m^i = \tilde{x}_m^{i-1} + \mathrm{d}\tilde{x}_m , \tag{33a}$$

$$\tilde{y}_m^i = \tilde{y}_m^{i-1} + \mathrm{d}\tilde{y}_m . \tag{33b}$$

$$\tilde{p}_m^i = \tilde{p}_m^{i-1} + \mathrm{d}\tilde{p}_m . \tag{33c}$$

At the iteration $I_1 + 1$ any given initial value for transmit power $\tilde{p}_m^{I_1} \in [\underline{P}, \bar{P}]$ is acceptable. We choose $\tilde{p}_m^{I_1} = \frac{1}{2}(\underline{P} + \bar{P}) , \forall m \in \mathbb{I}_M$, $M = N$. Indeed, the power until I_1^{th} iteration is not updated and we assume $\tilde{p}_m^i = \frac{1}{2}(\underline{P} + \bar{P}) , \forall i \in [1, I_1]$, $\forall m \in \mathbb{I}_M$, $M = NG^2$. Such an approach works very well if the transmit power of all targets are known, i.e., $\bar{P} = \underline{P}$. It, nonetheless, fails to deliver a good solution for $N > 2$ targets. We have observed this from extensive simulations.

B. ℓ_1-localization with unknown transmit power

We know that having devised the variables $\tilde{d}p_m$ enables updating the value of transmit power, when it is unknown. Therefore, we combine the objective functions of the optimization problems (23) and (31) and come up with the following convex problem

$$\min_{\substack{\tilde{s}_m,\mathrm{d}\tilde{x}_m,\mathrm{d}\tilde{y}_m,\\ \mathrm{d}\tilde{p}_m\,,m\in\mathbb{I}_M}} \sum_{k\in\mathbb{I}_K}\Big[\mu\big(\sum_{k'\in\mathbb{I}_K}\psi_{kk'}r_{k'}-\sum_{m\in\mathbb{I}_M}q_{km}\,\tilde{s}_m\big)^2 + \big(f_k^{i-1}+\sum_{m\in\mathbb{I}_M}a_{km}^{i-1}\mathrm{d}\tilde{x}_m+b_{km}^{i-1}\mathrm{d}\tilde{y}_m+c_{km}^{i-1}\mathrm{d}\tilde{p}_m\big)^2\Big] \tag{34a}$$

$$\text{s.t.}\;\; \tilde{s}_m, \mathrm{d}\tilde{x}_m, \mathrm{d}\tilde{y}_m, \mathrm{d}\tilde{p}_m \in \mathbb{R}\,, \tag{34b}$$

$$\underline{P}-\tilde{p}_m^{i-1}\le \mathrm{d}\tilde{p}_m \le \bar{P}-\tilde{p}_m^{i-1}\,, \tag{34c}$$

$$-\delta \le \mathrm{d}\tilde{x}_m \le \delta\,, \tag{34d}$$

$$-\delta \le \mathrm{d}\tilde{y}_m \le \delta\,, \tag{34e}$$

$$0 \le \tilde{s}_m \le 1\,, \tag{34f}$$

$$\sum_{m\in\mathbb{I}_M}\tilde{s}_m = N\,, \tag{34g}$$

where $\psi_{kk'}$ and q_{km} are the entries of the matrices $\mathbf{\Psi}$ and $\mathbf{Q}$, defined in (21) and (23) at the i^{th} iteration. The parameter $\mu = 1$ until iteration I_1 and then is set to 0 from the iteration $I_1 + 1$ onward. That means from $I_1 + 1$ the variable $\tilde{s}_m$ is of no importance to the localization algorithm. On the other hand, to make use of the variable $\tilde{s}_m$ we apply the averaging rule until iteration I_1

$$\hat{x}_n = \frac{\sum\limits_{m\in\Pi_n} s_m^\star(\tilde{x}_m^{i-1}+\mathrm{d}\tilde{x}_m^\star)}{\sum\limits_{m\in\Pi_n} s_m^\star}\,, \tag{35a}$$

$$\hat{y}_n = \frac{\sum\limits_{m\in\Pi_n} s_m^\star(\tilde{y}_m^{i-1}+\mathrm{d}\tilde{y}_m^\star)}{\sum\limits_{m\in\Pi_n} s_m^\star}\,, \tag{35b}$$

$$\hat{p}_n = \frac{\sum\limits_{m\in\Pi_n} s_m^\star(\tilde{p}_m^{i-1}+\mathrm{d}\tilde{p}_m^\star)}{\sum\limits_{m\in\Pi_n} s_m^\star}\,, \tag{35c}$$

where the partitions $\Pi_1, \cdots, \Pi_N$ can be found by k-means method, i.e., similar to (24). Then, the grid will be updated using (26) by generating a sub-grid of granularity G centered at each $(\hat{x}_n, \hat{y}_n)$. Consequently, the transmit power $\hat{p}_n$ will be assigned to each GP of the relevant sub-grid:

$$\mathcal{P}(N,G) = \bigcup_{n\in\mathbb{I}_N}\hat{p}_n\otimes\mathbf{1}_{G^2}\,, \tag{36}$$

where $\mathbf{1}_{G^2}$ is the all-ones vector of size G^2. Indeed, the set $\mathcal{P}(N, G)$, whose cardinality is NG^2, is the set of transmit power of all the grid points. The Alg. 1 summarizes the idea of the ℓ_1-localization of ours. As we will see in Sec. V, the algorithm has a very good performance for the case of unknown transmit power.

Algorithm 1 ℓ_1-localization heuristic for the joint estimate of the transmit power and location of multiple targets

initialization:
- set the grid granularity $G \in \mathbb{N}$
- set the area width $2w \in \mathbb{R}_+$
- $\delta \leftarrow \frac{w}{4(G-1)}$
- let $\hat{p}_n = \frac{1}{2}(\underline{P} + \bar{P})$ and $(\hat{x}_n, \hat{y}_n) = (0, 0)$, $\forall n \in \mathbb{I}_N$
- let $M = NG^2$ be the number of GPs
- let $M_0 = G^2$ to build the set $\bar{\mathbb{I}}_{M_0}$ and the partitions $\Pi_1, \cdots, \Pi_N$
- set the number of iterations $I_1, I_2 \in \mathbb{N}$
- $\mu \leftarrow 1$

for $i \leftarrow 1$ to I_1 **do**
  define $\mathcal{G}(N)$ using (26) and $\mathcal{P}(N, G)$ using (36)
  let $(\tilde{x}_m^{i-1}, \tilde{y}_m^{i-1}) \in \mathcal{G}(N)$, $\forall m \in \mathbb{I}_M$
  let $\tilde{p}_m^{i-1} \in \mathcal{P}(N, G)$, $\forall m \in \mathbb{I}_M$
  find optimal values $\tilde{s}_m^\star$, $\mathrm{d}\tilde{x}_m^\star$, $\mathrm{d}\tilde{y}_m^\star$, $\mathrm{d}\tilde{p}_m^\star$ using (34)
  calculate the estimate points $\hat{p}_n$, $\hat{x}_n$ and $\hat{y}_n$ using (35)
end for
- $M \leftarrow N$
- $\delta \leftarrow \frac{w}{G-1}$
- $\mu \leftarrow 0$

for $i \leftarrow I_1 + 1$ to $I_1 + I_2$ **do**
  define $\mathcal{G}(N)$ using (32) and $\mathcal{P}(N, 1)$ using (36)
  let $(\tilde{x}_m^{i-1}, \tilde{y}_m^{i-1}) \in \mathcal{G}(N)$, $\forall m \in \mathbb{I}_M$
  let $\tilde{p}_m^{i-1} \in \mathcal{P}(N, 1)$, $\forall m \in \mathbb{I}_M$
  find optimal values $\mathrm{d}\tilde{x}_m^\star$, $\mathrm{d}\tilde{y}_m^\star$, $\mathrm{d}\tilde{p}_m^\star$ using (34)
  $\hat{x}_n \leftarrow \tilde{x}_m^{i-1} + \mathrm{d}\tilde{x}_m^\star$, $\forall n \in \mathbb{I}_N, m = n$
  $\hat{y}_n \leftarrow \tilde{y}_m^{i-1} + \mathrm{d}\tilde{y}_m^\star$, $\forall n \in \mathbb{I}_N, m = n$
  $\hat{p}_n \leftarrow \tilde{p}_m^{i-1} + \mathrm{d}\tilde{p}_m^\star$, $\forall n \in \mathbb{I}_N, m = n$
end for
$\mathcal{X} \coloneqq \{(\hat{x}_n, \hat{y}_n, \hat{p}_n)\,|\,\forall n \in \mathbb{I}_N\}$
return $\mathcal{X}$

V. SIMULATIONS

Since there is no work with the same assumptions as ours, i.e., multi co-channel targets, we cannot compare our results with any other works, unfortunately. The only work, except for our previous papers [6], [7], that has similar assumptions is [5] which deal with a fingerprinting problem. Therefore, a fair comparison with its results is not straightforward. In what follows we evaluate the performance of the proposed ℓ_1-localization of this paper with the ℓ_0-localization [7]. The evaluations are done by means of computer simulations.

In the simulation setup $\bar{P} = 1$, $\underline{P} = 0.5$ and $w = 1$Km are chosen. The results are the outcome of $J = 5000$ simulation realizations, in each of which the position of sensors and realization of ζ_{kn}s are random, while the transmit power and position of targets are always the same. Let the estimated position of the n^{th} target at j^{th} realization be denoted by $(\hat{x}_n^j, \hat{y}_n^j)$. Then, *positioning root mean square error* (PRMSE) in meters is defined by, [5]:

$$\overline{\delta} = \sqrt{\frac{1}{JN}\sum_{j=1}^{J}\sum_{n=1}^{N}\Big(\hat{x}_n^j - x_n\Big)^2 + \Big(\hat{y}_n^j - y_n\Big)^2}, \tag{37}$$

Let the maximum positioning error at j^{th} iteration, i.e.,

$$\delta_{\max}^j \coloneqq \max_{n\in\mathbb{I}_N}\sqrt{\Big(\hat{x}_n^j - x_n\Big)^2 + \Big(\hat{y}_n^j - y_n\Big)^2}\,, \tag{38}$$

be a sample drawn from the distribution of a random variable, e.g., Δ. Then, the error function

$$P_d \coloneqq \Pr\left(\Delta > d\right) = 1 - F_{\Delta}(d)\,, \tag{39}$$

stands for the probability that at least one of the targets is localized with an error of more than d meters. Note, F_{Δ} is the empirical cdf of the error Δ. Similarly, the *root mean square error* (RMSE) of the transmit power is defined by

$$\overline{\rho} = \sqrt{\frac{1}{NJ}\sum_{j=1}^{J}\sum_{n\in\mathbb{I}_N}\rho_n^j}\,, \tag{40}$$

where $\rho_n^j \coloneqq (p_n - \hat{p}_n^j)^2$ is the square error of the estimated power value of n^{th} target at j^{th} realization.

The simulation results of the proposed ℓ_1-localization are given in Fig. 3 and Fig. 4 for $N = 2$ and $N = 3$ for different values of K and σ. The parameter σ represents the strength of shadowing. The figure also compares the performance of the algorithm with the method of combinatorial (ℓ_0-localization) of [7]. In the legend of the figures, the values of $\overline{\delta}$ and $\overline{\rho}$ are shown.

From the figures, we observe that $P_d \to 0, \forall d \geq 10{-}6$ in the case of $N = 2$ and $\gamma \to \infty$. Furthermore, for $\gamma = 40$ the positioning error of all the targets is very unlikely to be more than 10m. We see that increasing the number of sensors to $K = 50$ can improve the localization quality by a factor of 10. This means the $\Pr(\Delta > 1\text{m}) \to 0$ and PRMSE decreases down to 7 meters in an area of 4km^2. In general, the PRMSE ranges from 7 to 37 meters, in contrast to the one of ℓ_1-localization which can be up to 143.8 meters. One example is for $K = 10$ and $\gamma \to \infty$, this value has reduced from 143.8 to 17.3 meters by the method of ℓ_1-localization.

Unfortunately, due to the complexity of the ℓ_0-localization, we cannot increase the values of K or G to control that the gap between curves and $P_d = 0$ closes or not. As we see ℓ_0-localization never gets close to $P_d = 0$.

In the strong shadowing conditions, i.e., higher values of σ, we could deploy more SNs to make the estimation more reliable. This is hopefully viable since RSS-based localization requires inexpensive and not sophisticated sensors, on the one hand. On the other hand, the proposed ℓ_1-localization method has a low-complexity, i.e., $\mathcal{O}(N^3)$ as proved in Sec. VI, and can solve the problem for higher values of K, efficiently. Simulation shows that for the case $\gamma = 40$, $N = 2$, increasing K from 10 to 20 and 50 decreases the probability $\Pr(\Delta > 1\text{m})$ from 66% to less than, respectively, 22% and 0.1%.

We also see from Fig. 4, in case 3 targets the performance degrades, especially for $K = 10$. The reason is that the number of variables increases as the number of targets increases. Consequently, more equations, i.e., more number of sensors are needed. We see that for $\sigma = 40$, $K = 50$ the error probabilities $\Pr(\Delta > 1\text{m}) = 8\%$ and $\Pr(\Delta > 10\text{m}) < 2\%$ are achieved by ℓ_1-localization. In contrast to ℓ_1-localization, we see ℓ_0-localization for $N = 3$ becomes unreliable.

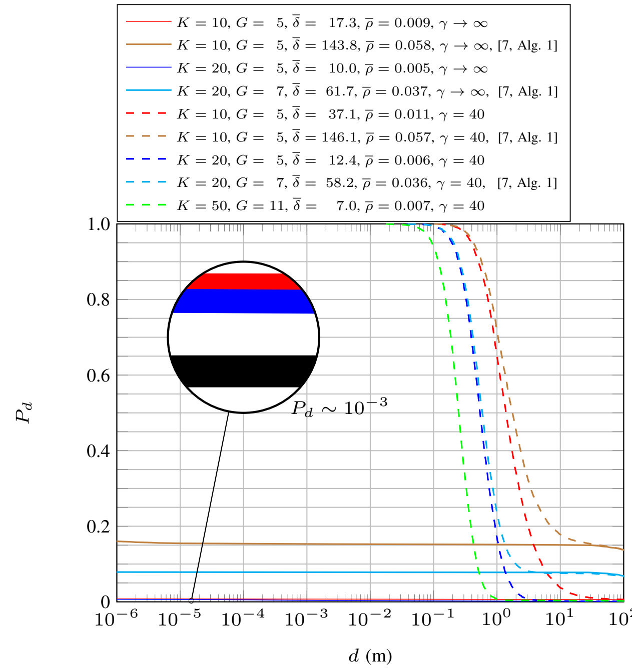

Fig. 3: The error probability P_d against positioning error d for $N = 2$ targets achieved by Alg. 1 and the ℓ_0-localization of [7]. The transmit power of targets are unknown and parameters $I_1 = 8$ and $I_2 = 12$ have been chosen. The values of $\overline{\delta}$ and $\overline{\rho}$ are shown in the legend.

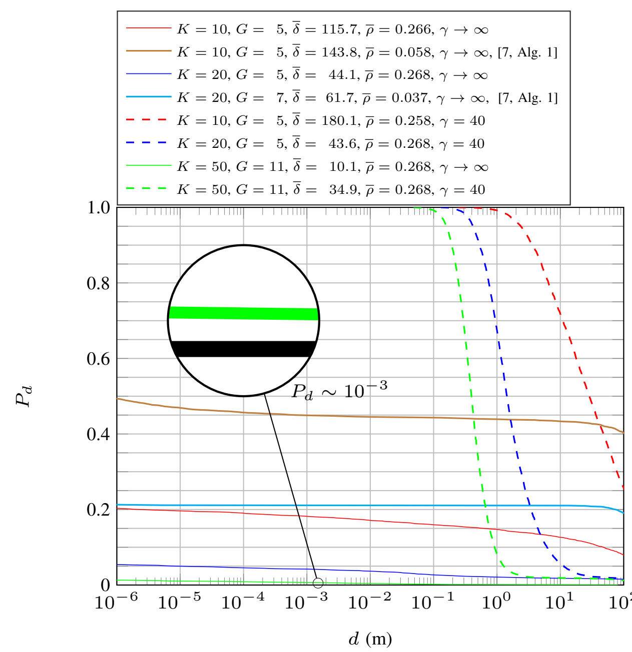

Fig. 4: The error probability P_d against positioning error d for $N = 3$ targets achieved by Alg. 1 and the ℓ_0-localization of [7]. The transmit power of targets are unknown and parameters $I_1 = 8$ and $I_2 = 12$ have been chosen. The values of $\overline{\delta}$ and $\overline{\rho}$ are shown in the legend.

VI. Complexity Analysis

Alg. 1 consists of three main stages as follows:

1) from iteration 1 to I_1: the problem (34) which belongs to the family of QP. Such problems are known to have a complexity of $\mathcal{O}(n^3)$, where n is the number of variables. In this problem $n = 4NG^2$ and thus the problem in all the I_1 iterations has a complexity of $\mathcal{O}(64I_1N^3G^6)$.
2) from iteration 1 to I_1: solving the k-means clustering problem to find the sets $\Pi_1, \cdots, \Pi_N$ for the averaging rule (24).

   There exist numerous algorithms to solve the k-means clustering problem, among which the Lloyd's heuristic algorithm is the most famous one. Its complexity is $\mathcal{O}(2G^2Ni)$ in the case of *two dimensional* (2D) localization and $\mathcal{O}(3G^2Ni)$ in the case of *three dimensional* (3D) localization [23]. Here i denotes the number of iterations for convergence, which is often small if the data has a clustering structure. Therefore, Lloyd's algorithm is known to have practically linear complexity, and its worst-case complexity is superpolynomial [24].
3) from iteration $I_1 + 1$ to $I_1 + I_2$: the QP in (31) with $3N$ variables. Because $\mu = 0$ causes the objective function to become independent from the variables $\tilde{s}_m$. Thus, the problem (34) reduces to (31). Indeed, the three variables $\mathrm{d}\tilde{x}_m$, $\mathrm{d}\tilde{y}_m$, and $\mathrm{d}\tilde{p}_m$ for each of the N targets make for $3N$ variables in total. Hence, this stage of the algorithm imposes a complexity of $\mathcal{O}(27I_2N^3)$.

   As the k-means clustering algorithm is linear in the number of targets, its complexity can be neglected in compassion with the one QP. Therefore, the complexity of Alg. 1 is $\mathcal{O}(64I_1N^3G^6 + 27I_2N^3)$, which yields $\mathcal{O}(N^3)$, i.e., a cubic order of complexity in the number of targets. Therefore, the presented ℓ_1-localization algorithm in this paper, unlike ℓ_0-localization in [7], has a polynomial complexity.

References

[1] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal, "Locating the nodes: cooperative localization in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 54–69, July 2005.

[2] J. H. Lee and R. M. Buehrer, "Location estimation using differential RSS with spatially correlated shadowing," in *Proceedings of the 28th IEEE Conference on Global Telecommunications*, ser. GLOBECOM'09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 4613–4618.

[3] C. Feng, S. Valaee, and Z. Tan, "Multiple target localization using compressive sensing," in *GLOBECOM 2009 - 2009 IEEE Global Telecommunications Conference*, Nov 2009, pp. 1–6.

[4] C. Feng, W. S. A. Au, S. Valaee, and Z. Tan, "Received-signal-strength-based indoor positioning using compressive sensing," *IEEE Transactions on Mobile Computing*, vol. 11, no. 12, pp. 1983–1993, Dec 2012.

[5] H. Jamali-Rad, H. Ramezani, and G. Leus, "Sparsity-aware multi-source RSS localization," *Signal Process.*, vol. 101, pp. 174–191, Aug. 2014.

[6] E. Zandi and R. Mathar, "RSS-based location and transmit power estimation of multiple co-channel targets," in *15th International Symposium on Wireless Communication Systems (ISWCS'18)*, Lisbon, Portugal, Aug. 2018, pp. 1–6.

[7] E. Zandi and R. mathar, "RSS-based positioning of multiple co-channel targets with unknown transmit power in a log-normal shadowing scenario," in *2018 IEEE International Conference on Wireless for Space and Extreme Environments (WiSEE'18)*, Huntsville, AL, USA, Dec. 2018, pp. 1–6.

[8] H. Jamali-Rad and G. Leus, "Sparsity-aware multi-source tdoa localization," *IEEE Transactions on Signal Processing*, vol. 61, no. 19, pp. 4874–4887, Oct 2013.

[9] H. Shen, Z. Ding, S. Dasgupta, and C. Zhao, "Multiple source localization in wireless sensor networks based on time of arrival measurement," *IEEE Transactions on Signal Processing*, vol. 62, no. 8, pp. 1938–1949, April 2014.

[10] T. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2001.

[11] P. Cardieri and T. S. Rappaport, "Statistics of the sum of lognormal variables in wireless communications," in *VTC2000-Spring. 2000 IEEE 51st Vehicular Technology Conference Proceedings (Cat. No.00CH37026)*, vol. 3, May 2000, pp. 1823–1827 vol.3.

[12] G. L. Stüber, *Principles of Mobile Communication (2Nd Ed.)*. Norwell, MA, USA: Kluwer Academic Publishers, 2001.

[13] X. Li, "Collaborative localization with received-signal strength in wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 6, pp. 3807–3817, Nov 2007.

[14] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[15] F. Chan, Y. T. Chan, and R. Inkol, "Path loss exponent estimation and RSS localization using the linearizing variable constraint," in *MILCOM 2016 - 2016 IEEE Military Communications Conference*, Nov 2016, pp. 225–229.

[16] T. Gerkmann and R. C. Hendriks, "Unbiased mmse-based noise power estimation with low complexity and low tracking delay," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 4, pp. 1383–1393, May 2012.

[17] J. Vartiainen, H. Saarnisaari, J. J. Lehtomaki, and M. Juntti, "A blind signal localization and snr estimation method," in *MILCOM 2006 - 2006 IEEE Military Communications conference*, Oct 2006, pp. 1–7.

[18] L. Fenton, "The sum of log-normal probability distributions in scatter transmission systems," *IRE Transactions on Communications Systems*, vol. 8, no. 1, pp. 57–67, March 1960.

[19] C. C. Heyde, *On a Property of the Lognormal Distribution*. New York, NY: Springer New York, 2010, pp. 16–18.

[20] M. B. Hcine and R. Bouallegue, "Fitting the log skew normal to the sum of independent lognormals distribution," *CoRR*, vol. abs/1501.02344, 2015. [Online]. Available: http://arxiv.org/abs/1501.02344

[21] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[22] I. Gurobi Optimization, "Gurobi optimizer reference manual, version 7.0, copyright © 2017," http://www.gurobi.com.

[23] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[24] D. Arthur and S. Vassilvitskii, "How slow is the k-means method?" in *Proceedings of the Twenty-second Annual Symposium on Computational Geometry*, ser. SCG '06. New York, NY, USA: ACM, 2006, pp. 144–153.